

A MESTERSÉGES INTELLIGENCIA ÉS A KÖNYVTÁROSOK

WINKLER BEA

Állatorvostudományi Egyetem, Hutýra Ferenc Könyvtár, Levéltár és Múzeum,
könyvtárigazgató

TARTALMI ÖSSZEFOGLALÓ

A mesterséges intelligencia (MI) napjaink egyik legdinamikusabban fejlődő területe. Ismertsége és használata a mindennapokban folyamatosan nő. Egy Google-kereséssel az artificial intelligence kifejezésre több mint 5 millió, magyar megfelelőjére, a mesterséges intelligenciára is több mint 4 millió találatot kapunk.

Jelen kutatás középpontjában az MI könyvtári szakirodalombeli megjelenésének mennyiségi elemzése állt. Két nagy, könyvtári szakirodalommal foglalkozó adatbázisban végzett keresés biztosította a kutatás alapját, ezek a Library & Information Science Collection és a Library, Information Science & Technology Abstracts.

Az eredmények az összesen kicsit több mint 300 publikáció alapján azt mutatják, hogy napjainkban a teljes könyvtári szakirodalomnak csak csekély része foglalkozik az MI-vel, azonban a 2016-os évtől jelentős bővülés tapasztalható a megjelent cikkek száma terén.

Bevezetés

1982-ben a *Time* magazin *Az év emberének* választotta a személyi számítógépet, 2006-ban mindenki *Az év embere* lett, aki alakítja és formálja a digitális társadalmat. A XXI. században ugyan még nem választották meg *Az év emberének*, de mindenképp nagyratörő szereplő a mesterséges intelligencia (artificial intelligence, AI, a továbbiakban: MI).

Az MI kezdetei egészen az ókori filozófiáig nyúlnak vissza, a mai értelemben vett megvalósítására azonban a XX. századig kellett várni. Az elmúlt években egyre több területen jelent meg a mesterséges intelligencia, például a szakértői rendszerekben, a robottechnikaiban, az önvezető járművek terén, de feltűntek a chatbotok, go játékosok, személyi asszisztensek stb.

A mesterséges intelligencia önálló területként való megjelenésének az 1956-ban „*Dartmouthban tartott találkozót tekintik, ahol összehozták azokat az amerikai kutatókat, akik érdekeltek az automataelméletben, a neurális hálókból és az intelligencia kutatásában,*” amint azt Russell és Norvig¹ is olvashatjuk.

Napjainkra számtalan feladatra hoztak létre specializált mesterséges intelligenciákat, amelyek egy-egy területen kiemelkedően teljesítenek, emellett gondolatilag már létrejött a mindenre alkalmas *superintelligencia*² is. A hétköznapi életnek ezek egyre inkább részévé válnak, még ha az átlagember sok esetben nem is gondolja, hogy mennyire.

Magam azért választottam ezt a témát, mert azt gondolom, hogy a könyvtári munkának számtalan olyan aspektusa van, ahol nagy sikerrel lehetne alkalmazni a mesterséges intelligenciát. Az így felszabaduló könyvtárosi kapacitással pedig bővíthetnénk a kreativitást és érzékenységet igénylő területeken a jelenlétünket.

A tanulmány célja annak feltérképezése, hogy két nagy könyvtári szakirodalmi adatbázisban hogyan jelent meg a mesterséges intelligencia és a könyvtár kapcsolata az elmúlt évtizedekben. A kutatás nem az egyes cikkek tartalmára fókuszált, hanem a megtalált cikkek időbeli mennyiségi változására, illetve a tárgyszavak mennyiségi elemzésére.

Mi a mesterséges intelligencia és hogyan kapcsolódik ez a könyvtárakhoz?

Russell és Norvig *Mesterséges intelligencia – Modern megközelítésben*¹ című könyvükben a fogalom meghatározásakor több csoportra bontották a lehetséges/szükséges feltételeket: Turing teszt, kognitív modellezés, a gondolkodás törvénye¹ stb. Ezek azonban így leírva az átlagember számára nehezen feldolgozható fogalmak.

A legtöbb esetben MI alatt az olyan „számítógépes rendszerek elméletét és fejlődését értjük, amelyek képesek olyan feladatokat végezni, amelyek általában emberi intelligenciát igényelnek, például a vizuális felfogást, a beszédfelismerést, a döntéshozatalt, a nyelvek közötti fordítást,” amint azt a Oxford Reference-ben³ is olvashatjuk.

A könyvtárak fejlődése mindig is szorosan követte az információ tárolásának és szolgáltatásának technológiai fejlődését. Napjainkban már számtalan területen folynak kutatások, amelyek célja a mesterséges intelligencia könyvtári alkalmazási területeinek meghatározása, kipróbálása.

A mesterséges intelligencia alkalmazásának lehetőségei a könyvtárakban először a szakértői rendszerekkel és adatbázisokkal jelentek meg. Az 1980-as évek vége felé a publikációk már könyvtári robotokról, számítógép által végzett referenz szolgáltatásról szóltak.

Az új évezredre a szakirodalomban megjelentek a virtuális és digitális könyvtárak, az intelligens tanulni képes könyvtárosi feladatokat ellátó robotok, a kiterjesztett valóság, amely akár az olvasás népszerűsítésében is tényező lehet egyes fejlesztők szerint.

A kutatás tárgya és módszerei

Kutatásom a következő előfeltevésekkel indult:

1. A mesterséges intelligenciával foglalkozó publikációk száma folyamatosan növekszik mind a kiindulási év cikkszámához képest, mind pedig az aktuális évet/ időszakot megelőző évhez/ időszakhoz képest.
2. A mesterséges intelligenciával foglalkozó publikációk, amelyek az „artificial intelligence” tárgyszót tartalmazzák, több ezren lehetnek a könyvtártudományal foglalkozó adatbázisok mindegyikében.

3. A mesterséges intelligenciával foglalkozó publikációk, amelyek a keresések során használt szakadatbázisokban hasonló keresések találatai voltak, hasonlóak lesznek.
4. A mesterséges intelligenciával foglalkozó publikációk, a két adatbázisban évenként vizsgálva azonos mennyiségűek lesznek.

Szerettem volna, ha minél szélesebb körű a kutatás, ezért a következő adatbázisokban próbakereséseket végeztem:

- Library & Information Science Collection
- Library, Information Science & Technology Abstracts
- Scopus
- Web of Science
- illetve az Ebsco Discovery Service segítségével több mint 80 adatbázisban (Europeana, BASE, Academic Search Complete, ScienceDirect, DOAJ stb.).

Az elsődleges keresőkifejezés kezdettől fogva az *artificial intelligence* volt, amely már a kutatási projekt címében is szerepel. A próbakeresések során megállapítottam, hogy a sok esetben szinte szinonimaként alkalmazott *machine learning*, *robotics*, *thinking computer*, *expert system* stb. szavak, kifejezések közül könyvtári témára vetítve a legrelevánsabb találatokat az *expert system* adta. Ezért a kutatás során az *artificial intelligence* és az *expert system* kifejezések lettek a keresés alapjai a téma meghatározó fogalmaiként. A szűk meghatározás indoka az volt, hogy a mennyiségi elemzések elvégzésére fókuszáló kutatás során minél pontosabb találatokkal akartam dolgozni. Ekkor merült fel lehetőségként, hogy a találati listák utólagos kézi szűrésen essenek át. A kézi szűrés előnye, hogy a keresés során sokféle és tág értelmű kifejezéssel lehetne keresni. E megoldás egyik jelentős hátránya viszont az lenne, hogy a találati listákat nagyon erősen befolyásolná a személyes preferencia.

A fenti megoldás helyett, ezért a keresés során alkalmazott és előre meghatározott szisztéma alapján elvégzett szűrés mellett döntöttem, amely segített kizárni a nem könyvtárakkal kapcsolatos találatokat. A kutatásban a fenti megfontolásból szűkítésként a következő kifejezéseket alkalmaztam: *library*, *libraries* és *librarians*.

A próbakeresések eredményei alapján meghatároztam az adatbázisok körét: Library & Information Science Collection (továbbiakban LISC) és Library, Information Science & Technology Abstracts (továbbiakban LISTA).

A *Library & Information Science Collection* a Proquest adatbázisa, a könyvtártudomány egyik vezető indexe, mely számtalan teljes szöveges hozzáférést is tartalmaz. Feldolgozza a szakterület tudományos közleményeit, konferencia anyagokat és egyéb, a kutatók és a szakemberek számára fontos tartalmakat. Húsz különböző nyelven, 1969-től napjainkig található meg benne a bibliográfiai leírások, illetve a publikációk teljes szövegei is.

A *Library, Information Science & Technology Abstracts* az EBSCO adatbázisa, az 1960-as évek közepétől napjainkig lefedi a könyvtártudomány magfolyóiratait, a legfontosabb monográfiákat és kutatási jelentéseket. Anyagai egyes része nem pusztán index, hanem a teljes szövegek is azonnal elérhetőek.

A LISC- és a LISTA-adatbázisban a kiválasztott kifejezések pontos egyezéssel, tárgyszóként kerültek lekeresésre, amelyet a találati listák alapján három kizárással egészítettem ki:

su(„artificial intelligence” **OR** „expert system”) **AND su**(„library” **OR**
„libraries” **OR** „librarians”)

NOT su(„Peptide Library” **OR** „Small Molecule Libraries” **OR** „Gene Library”)

A jelen kutatás alapjául szolgáló adatbázis-keresések elvégzésének utolsó időpontja 2017. október 17-e és 20-a között volt.

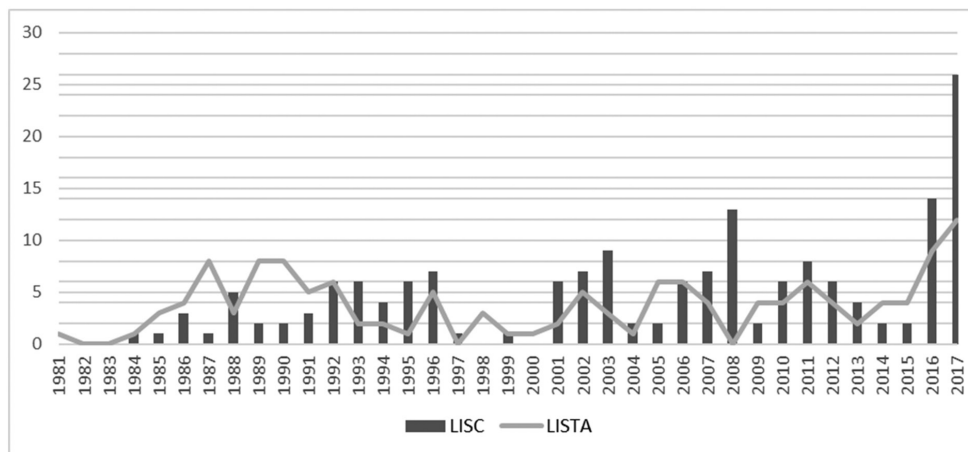
Eredmények

A LISC-adatbázisban a keresés találati listáján 218 publikáció szerepel, amely letöltésnél 172 lesz, majd kézi duplum szűrés után 171-re csökken. Nem túl meglepő, hogy a kapott publikációk 95 százaléka angol nyelvű és az első nem angol nyelvű publikáció a találati listán csak 1990-ből való.

A LISTA-adatbázisban 138 tétel a találati lista. Ennek nagy része szintén angol nyelvű, csak elvétve található egy-egy horvát, olasz stb. nyelvű, sőt van egy magyar nyelvű is a Tudományos és Műszaki Tájékoztatásból 2011-ből.

A II. előfeltevés a találati listák eredményei alapján nem állja meg a helyét, a vártnál sokkal kevesebb publikációt találtam, mivel összesen a két adatbázisban alig több mint 300 cím szerepel.

A publikációk mennyiségére vonatkozó I. előfeltevés az volt, hogy 1981-től minden évben több publikáció jelent meg, mint az azt megelőző évben, amelyet az 1. ábrán látható adatok nem erősítenek meg.



1. ábra: A találatok darabszáma 1981 és 2017 között évenként vizsgálva a LISC és LISTA adatbázis esetén

Az 1. ábra bemutatja a LISC- és LISTA-adatbázisok találati listáinak évekre bontott adatait. Az ábra alapján jól látható, hogy a cikkek évenkénti száma a két adatbázisban egymástól függetlenül alakult. A két lista évenkénti cikkszámának korrelációs együtthatója 0,515, ami megerősíti a látottakat, hogy a két adatsor alakulása között nincs kapcsolat. A IV. feltevést tehát nem igazolják az adatok.

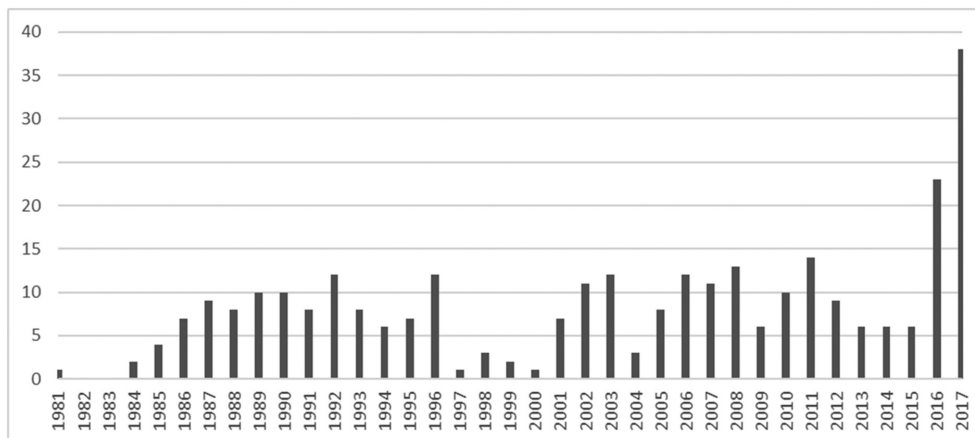
A kutatott téma könyvtári szakirodalmon belüli nagyságának felmérésére lefuttattam egy keresést a következő formában: **su**(„library” **OR** „libraries” **OR** „librarians”) **NOT su**(„Peptide Library” **OR** „Small Molecule Libraries” **OR** „Gene Library”).

A LISTA-adatbázisban 208 ezer, a LISC-adatbázisban több mint 258 ezer találatot kaptam. A két adatbázis találati listáján szereplő cikkek darabszáma alapján elmondható, hogy napjainkban nem a mesterséges intelligenciával foglalkozó kutatások állnak a könyvtári kutatások fókuszában.

Mivel két alapvetően hasonló profilú adatbázisról volt szó, összehasonlításra került a két találati lista, amelyből kiderült, hogy mindösszesen három olyan cikk van, amely mindkét listán szerepel. A két gyökeresen eltérő találati lista egyik oka lehetne, hogy a két adatbázis más anyagokat dolgoz fel. Az elérhető leírások alapján a LISC⁴, ami közel 500 címet tartalmaz és a LISTA⁵, ami 830 címet tartalmaz, több mint 200 közös címet dolgoz fel, így ez nem ad egyértelmű magyarázatot a kapott eredményekre.

Egy másik, a fentieket kiegészítő lehetőség, hogy mivel a keresések során tárgyszavakra leszűkítve kerestem, a két adatbázis eltérő tárgyszavazási metódusa miatt tér el ennyire a két találati lista. Azt megállapítottam, hogy az elvégzett keresés esetében a két adatbázis találati listája a téma szakirodalmára tekintettel kiegészíti egymást.

Az előző megállapítás eredményeképpen az évenként megjelent cikkek darabszámának további elemzéséhez a két listát egyesítettem és a duplumokat kiszűrtem.



2. ábra: A találatok évenkénti darabszáma a LISC- és a LISTA-adatbázis találatainak összevonása után 1981 és 2017 november között

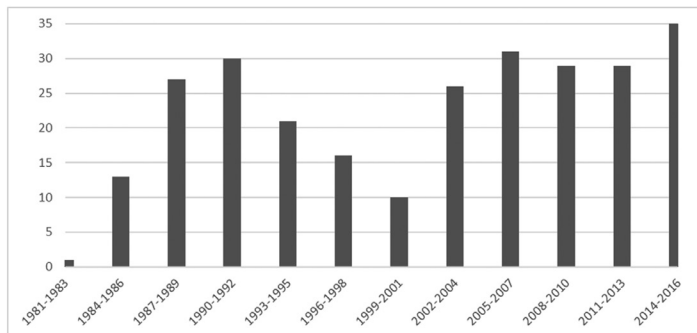
A LISC és LISTA közös találati lista évenkénti adatai láthatók a 2. ábrán. Az évek és az adott évi cikkek darabszáma között nincs egyenletes emelkedő trend továbbra sem. Inkább egyfajta hullám figyelhető meg, amelyet a 2016-os és a 2017-es év tör meg jelentősen. A 2. ábra alapján nehéz meghatározni a jövőre vonatkozó előrejelzést.

A mesterséges intelligencia kutatása általában sem egy folyamatosan emelkedő kutatási ívet írt le az elmúlt hatvan esztendőben. Az MI történetével is foglalkozók egyfajta hullámzó eseménysort határoztak meg, amely a következők nyomán alakult ki: a téma kutatásának megindulása, majd az első eredmények megjelenése, publikussá válása, a támogatás megindulása, gyors sikerek; hatalmas támogatások és velük együtt a túlzott elvárások kialakulása; a túlzott elvárások nem teljesülnek, a befektetők csalódnak, a támogatások csökkennek; a téma „jégelése”, az ún. „mesterséges intelligencia-tél” („MI-tél”) kialakulása, amikor a kutatások inkább a háttérben folynak, kevesebb anyagi támogatással.

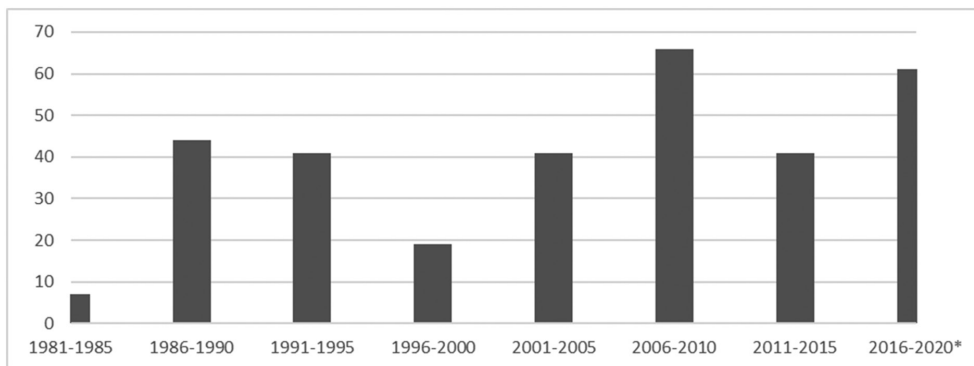
A mesterséges intelligenciával kapcsolatos, a könyvtári területéről szóló publikációk jelen kutatásom eredményei alapján jóval később jelentek meg, mint a modern MI-kutatások indulása (1950-es évek). Az 1980-as években megjelent első néhány tucat közlemény után az 1990-es évek közepén a könyvtári területen is megjelent az „MI-tél”, visszaesett a publikált kutatások száma, amely az új évezred hajnalán ért véget. A 2010-es évek elején ismét csökkent a publikációk száma, hogy aztán minden eddiginél nagyobb emelkedéssel térjen vissza napjainkban.

Az évenkénti cikkszámok hosszabb távon valószínűleg tovább fognak emelkedni, azonban ennek volumenét ma még nehezen lehetne bizonyíthatóan megjósolni, mert éppúgy elképzelhető lineáris emelkedés, mint exponenciális. Ahogy Ray Kurtzweil is írja, *az exponenciális emelkedés csalogója: lassan, gyakorlatilag észrevétlenül kezdődik, de a görbe térdén túl robbanásszerűvé és mélyen átalakítóvá válik*⁶.

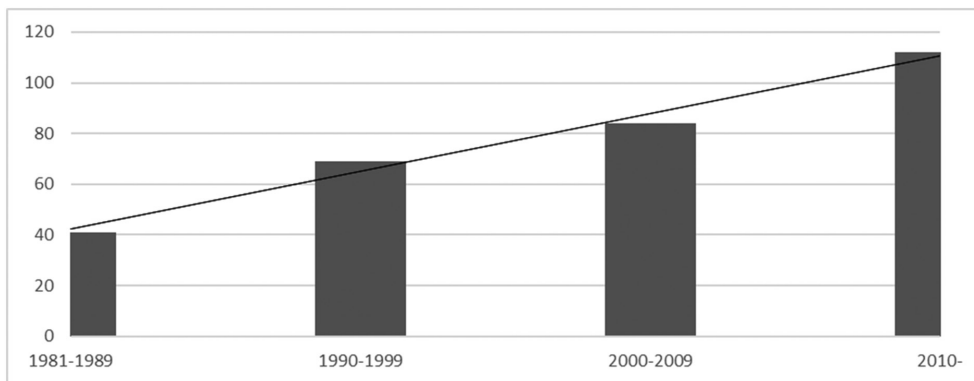
A LISC és LISTA egyesített találati lista adatait az évenkénti elemzés után 3 éves, 5 éves és 10 éves időszakokra bontva is összesítettem. A cél az volt, hogy a hosszabb időszakok esetén jobban kimutathatók legyenek az esetleges rövid intervallum miatt rejtve maradó trendek, hiszen a teljes kutatás sem ölel fel negyven évet. (3. ábra)



3. ábra: A találatok darabszámainak 3 évenkénti összegzett darabszámai a LISC- és a LISTA-adatbázis találatainak összevonása után 1981 és 2017 között



4. ábra: A találatok darabszámainak 5 évenként összegzett darabszámai a LIS- és a LISTA-adatbázis találatainak összehasonlítása után 1981 és 2017 között



5. ábra: A találatok darabszámainak 10 évenként összegzett darabszámai a LIS- és a LISTA-adatbázis találatainak összehasonlítása után 1981 és 2017 között, lineáris trendvonalal

A 4. és az 5. ábra adatainak vizsgálatakor figyelembe kell venni, hogy az utolsó adatsor mindkét esetben hiányos még, mert jelentősen a meghatározott időhatár vége előtt vagyunk. A 4. ábrán látható 2016–2020-as időszaknak több mint a fele hiányzik, az 5. ábra estén pedig a 2010–2019 közötti időszaknak több mint az ötöde. A 3. és 4. ábra adatain a 2. ábrán már látott cikkszám változási hullámot láthatjuk, benne az 1997–2000 közötti cikkszám visszaeséssel. Az 5. ábra igazolja az I. előfeltevés egyik részét, amely szerint a cikkszám növekedés a vizsgált időszakokban a kezdetek óta folyamatosan emelkedik az adott időszakot megelőző időszakhoz képest. Az időszakok és a megjelent cikkek darabszáma közötti korrelációs együttható 0,99, ami nagyon erős kapcsolatra utal.

Kutatásom következő szakasza a tárgyszavakra koncentrált. A letöltött találatok esetén a tárgyszavak a LISTA: Subject és a LIS- SubjectTerms mezőjében szerepeltek. A LIS (171 találat) találati listájához 1483, a LISTA-éhoz (138 találat) pedig 773 tárgyszó tartozott.

1. táblázat

A találati listákon leggyakrabban előforduló tárgyszavak (min. 10 cikk esetén használt kifejezések)

Library, Information Science & Technology Abstracts		Library & Information Science Collection	
tárgyszó	előfordulás (db)	tárgyszó	előfordulás (db)
Artificial intelligence	138	Artificial intelligence	124
Expert systems	40	Expert systems	12
Libraries – Automation	25	Index Medicus	28
Digital libraries	12	Digital libraries	26
Information science	12	Academic libraries	15
Library science	11	Studies	15
Academic libraries	11	Algorithms	13
Library users	10	Computers	12
		Automation	10
		Data mining	10
		Library science	10

Az 1. táblázatban is jól látható, hogy a kereső kifejezésként használt *artificial intelligence* nem meglepő módon mind a két esetben nagy súllyal van jelen. A LISC-lista legnépszerűbb tárgyszavai között jelentős számban találhatók informatikai kifejezések, mint az algoritmus, számítógépek, adatbányászat és az automatizálás. A LISTA esetén inkább a klasszikus könyvtári fogalmak és kifejezések domináltak: digitális könyvtár, könyvtártudomány, felsőoktatási könyvtárak, könyvtárhasználók stb., illetve itt is megjelent az automatizálás.

A tárgyszavak között a LISC-adatbázisban jelentős arányban megjelent az *index medicus*, amelyet nem zártam ki a keresés során azért, mert a találatok nagy része releváns, könyvtári találat volt.

A megvizsgált tárgyszavak alapján a mesterséges intelligencia és a könyvtárak témakörében az elmúlt 36 évben megjelent cikkek jelentős része technológiai jellegű fejlesztéseket mutat be. A kutatások többsége újabb és újabb technológiai megoldásokkal foglalkozik, amelyek automatizálhatnak vagy átalakíthatnak munkafolyamatokat, szolgáltatásokat.

Összefoglalás

Kutatásom alapját a LISC- és a LISTA-adatbázis több mint 300 publikációt tartalmazó listája alkotja. Az adatok a mesterséges intelligencia (MI) és a könyvtárak kapcsolatának 1981-től napjainkig tartó közös pontjait írják le.

A kutatás alapját képező találati lista rövidebbé okán következtethetünk arra, hogy a teljes könyvtári szakirodalomra vetítve az MI-vel foglalkozó rész csupán egy szűk kör

érdeklődésére tartott számot napjainkig. A megjelent publikációk száma tízéves időszakokat vizsgálva, folyamatos emelkedést mutat az idő előrehaladtával. A 2016-os és 2017-es év publikációs szám-bővülése előre vetíti, hogy a téma valószínűleg a közeljövőben fog igazán kibontakozni.

A találati listához kapcsolódó tárgyszavak alapján elmondható, hogy a mesterséges intelligencia könyvtárhasználókra, illetve a könyvtárosokra (mint emberekre) gyakorolt hatásával jelenleg még csak keveset foglalkoztak.

Kíváncsian tekintek a közeli jövőbe, hogy mely területeken fog a könyvtárak és könyvtárosok napi munkájában átalakulást hozni a mesterséges intelligencia elterjedése.

Jegyzetek

1. NORVIG, Peter – RUSSELL, Stuart J.: Mesterséges intelligencia – Modern megközelítésben. Budapest, Panem Kft., 2005.
2. BOSTROM, Nick: Szuperintelligencia. Budapest, Ad Astra, 2016. (e-könyv) [2018. január 14.]
3. Artificial intelligence. Forrás: <http://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960> [2018. január 9.]
4. Library & Information Science Collection – Title Lists System. Forrás: <http://tls.search.proquest.com/titlelist/ListForward?productId=1007519%3A1006545&productName=Library+%26+Information+Science+Collection&format=html&IDString=1007519%3A1006545> [2018. január 11.]
5. Library, Information Science & Technology Abstracts (LISTA) – Database Coverage List. Forrás: https://www.ebscohost.com/titleLists/lxh-coverage.htm?_ga=2.249990278.1866146030.1515687787-1854539136.1495695918 [2018. január 11.]
6. KURZWEIL, Ray: A szingularitás küszöbén. Budapest, Ad Astra, 2013. (e-könyv) [2018. január 14.]

Winkler Bea az Állatorvostudományi Egyetem, Hutýra Ferenc Könyvtár, Levéltár és Múzeum igazgatója. 2005 szeptemberétől a Szent István Egyetem Állatorvos-tudományi Könyvtárának könyvtárosa, majd igazgatóhelyettese és 2016 októberétől igazgatója. Az ELTE Irodalomtudományi Doktori Iskola Könyvtártudományi Doktori Programjának hallgatója. Fő érdeklődési területe a könyvtári területen alkalmazott informatikai megoldások és ezek speciális esete a mesterséges intelligencia.